Algorithm 1 Training Stage of EEG2Video Framework

Input: (1) training set $\mathcal{D}_{train} = \{x_i, v_i, d_i\}$, where x_i is EEG segments, v_i is video clips, d_i is the *fast/slow* label (2) stable diffusion model $T2I$, whose VAE encoder is $\mathcal{E}_{vae}$ (3) image caption model B

Output: (1) video diffusion model $T2V$, (2) Seq2Seq model $Seq2Seq$, (3) semantic predictor $\mathcal{P}_s$, (4) dynamic predictor $\mathcal{P}_d$

Initialize text prompts of training dataset $T = \{t_i\}$
for each $(v_i) \in \mathcal{D}_{train}$ **do**
    $v_i = \{f_1, f_2, \ldots, f_n\}$
    $t_i \leftarrow B(f_1)$
end for
Initialize latent vectors of all frames $L = \{z_i\}$
for each $(v_i) \in \mathcal{D}_{train}$ **do**
    $v_i = \{f_1, f_2, \ldots, f_n\}$
    $z_i = \{l_1, l_2, \ldots, l_n\}$
    for each $(f_j) \in v_i$ **do**
        $l_j \leftarrow VAE(f_j)$
    end for
end for
Fine-tune the $T2I$ with $\{v_i, t_i\}$ to obtain video diffusion model $T2V$
Train the Seq2Seq model $Seq2Seq$ with all $\{x_i, z_i\}$ using MSE loss
Train the semantic predictor $\mathcal{P}_s$ with all $\{x_i, t_i\}$ using MSE loss
Train the dynamic predictor $\mathcal{P}_d$ with all $\{x_i, d_i\}$ using Cross Entropy loss
return $T2V, Seq2Seq, \mathcal{P}_s, \mathcal{P}_d$

Algorithm 2 Inference Stage of EEG2Video Framework

Input: (1) EEG segments x_i from test set $\mathcal{D}_{test}$, (2) video diffusion model $T2V$, (3) Seq2Seq model $Seq2Seq$, (4) semantic predictor $\mathcal{P}_s$, (5) dynamic predictor $\mathcal{P}_d$

Output: reconstructed videos $\hat{v}_i$,

$\hat{z}_i \leftarrow Seq2Seq(x_i)$, $t_i \leftarrow \mathcal{P}_s(x_i)$, $d_i \leftarrow \mathcal{P}_d(x_i)$
Randomly sample $\epsilon_d = \{\epsilon_d^1, \epsilon_d^2, \ldots, \epsilon_d^n\}$, each $\epsilon_d^i \sim \mathcal{N}(0, 1)$
Randomly sample $\epsilon_s = \{\epsilon, \epsilon, \ldots, \epsilon\}$, where $\epsilon \sim \mathcal{N}(0, 1)$
$z_\mathcal{T} \leftarrow \sqrt{\overline{\alpha}_\mathcal{T}} \times \hat{z}_i + \sqrt{1 - \overline{\alpha}_\mathcal{T}} \times (\sqrt{\beta} \times \epsilon_s + \sqrt{1 - \beta} \times \epsilon_d)$, where β is determined by d_i
Generate videos $\hat{v}_i = T2V(z_\mathcal{T}, t_i)$
return $\hat{v}_i$